

RESPONDENT CUES THAT SURVEY QUESTIONS ARE IN DANGER OF BEING MISUNDERSTOOD¹

Jonathan Bloom, Dragon Systems
Michael F. Schober, New School for Social Research
Jonathan Bloom, Dragon Systems, 320 Nevada Street, Newton, MA 02460

Key words: Question comprehension, linguistic cues, pragmatic cues, misunderstanding

INTRODUCTION

In ordinary conversation, people who are uncertain about what their partners mean don't necessarily ask for clarification explicitly. Sometimes they wait to hear more, as if all were well, in hopes that something their partners say later will clarify what was meant. Whether it is because people want to avoid challenging their partners, because they want to avoid looking stupid, because they want to maintain the flow of the conversation, or because they honestly don't recognize that they haven't understood, people don't ask for clarification on many occasions when it would be rational to do so. For example, college students in one study rarely asked questions about math problems even when the problems contained anomalous and irrelevant information (Graesser & McMahan, 1993). In another study (Schober & Conrad, 1997), survey respondents answering factual questions misunderstood what the survey designers meant by questions as much as 73% of the time, but they asked for clarification only 38% of the time.

This poses a problem for survey researchers who want to be sure that respondents interpret questions as intended. The common wisdom is that interviewers should read questions exactly as worded, leaving the interpretation up to respondents so that interviewers won't bias the answers (e.g., Fowler & Mangione, 1990). But a common practice in survey organizations that consider themselves to follow standardization is for interviewers to read scripted definitions for key question concepts if respondents ask for them. (Technically, this practice isn't strictly standardized, in that some respondents get clarification and others don't, but it is nonetheless commonly considered to fall within the boundaries of standardized practice). The problem is that if respondents don't ask for clarification when they need it, how can interviewers provide clarification appropriately?

The problem is greater than it at first appears. In order to ask an appropriate clarification question, survey respondents must overcome more than just their fears of looking ignorant. They must recognize that their own way of conceiving of a question concept may not coincide with the survey designer's. They must also recognize that their personal circumstances warrant their pursuing clarification (see Schober, 1999; Schober & Conrad, 2000). Consider a respondent to the Current Point of Purchase Survey who is asked "Last year, did you purchase or have expenses for household furniture?"; and imagine that during the past year she bought a floor lamp. According to the official definitions of the sponsoring agency (the Bureau of Labor Statistics), the floor lamp shouldn't be counted as a furniture purchase. For the respondent to ask the right clarification question (something like "Do you count a floor lamp as a piece of household furniture?") she must recognize that her notion of furniture may differ from the survey designers'. She must also recognize that a floor lamp may be an instance that the survey designers might not count as a purchase of household furniture.

From our point of view, knowing when to ask for clarification and asking the right question are substantial metacognitive feats, and much remains unknown about people's metacognition. It certainly doesn't seem particularly surprising, given the complexity of what survey respondents are faced with, that they don't always ask for clarification when they ought to.

STUDY

The aim of the current study was to identify cues that survey researchers could exploit to determine whether respondents who don't explicitly ask for clarification are in danger of misunderstanding. Because there is substantial psycholinguistic and sociolinguistic evidence that people provide indirect evidence of their uncertainty in ordinary discourse, we focus on potential paralinguistic and pragmatic discourse cues. We focus on surveys of factual

¹ We thank Susan Brennan, Fred Conrad, and Bill Hirst for their comments, and Eduardo Vega for help with coding. This material is based upon work supported by the National Science Foundation under grants No. IRI-94-02167 and SBR-97-30140, and a dissertation fellowship from the New School for Social Research.

matters where we can assess whether respondents' interpretations match the survey designers' intentions (as embodied in the official definitions of survey concepts). We also focus on one particular kind of uncertainty that can arise in a survey situation: uncertainty about survey designers' definitions of survey concepts.²

From examining the discourse literature we identified several potential cues of respondent uncertainty. One possible set of cues consists of various sorts of *speech disfluencies*, which have been argued to provide evidence that speakers are having trouble in planning or executing their utterances. These include *ums* and *uhs*, too-long pauses in the wrong places, and self-corrections. In the current study we focused on the following disfluencies:

- (a) mid-clause or pre-utterance *pauses* lasting longer than one second (see Jefferson, 1989, for evidence that a one-second minimum is the reasonable lower bound). An example of a mid-clause pause of this sort would be found in a response to the household furniture question like "Yes I (1.2 second pause) did buy some furniture." An example of a pre-utterance pause would be "(2.3 second pause) Yes I did buy some furniture."
- (b) *Fillers* (also called *filled pauses*) like *um*, *uh*, and *mm*, as in a respondent's answer "Yes I did buy some um furniture" or "Uh yes I did buy some furniture."
- (c) *Repairs*, in which speakers correct what they say (see Bortfeld et al., under review; Levelt, 1989). In the answer "Y- y- yes I did buy a fl- some furniture" the respondent would have made several repairs.

Empirical evidence from ordinary (non-survey) discourse shows that these disfluencies can be correlated with speaker uncertainty. For example, people who are less certain about answers to trivia questions pause and produce fillers more often

²There are, of course, many other possible sources of respondent uncertainty in survey interviews (see Schaeffer, Maynard & Cradock, 1993). Respondents can be uncertain about what was said, because they did not hear the question clearly (did the interviewer say *farm* or *firm*?). Respondents may be uncertain about their own memory (did I buy milk last week or not?). Or they can be uncertain about the nature of the response task (should I count all instances or estimate?—see Conrad, 1999). Each of these other kinds of uncertainty may have its own sorts of cues.

(Smith & Clark, 1993). And listeners have been shown to judge a speaker's feeling of knowing accurately based on pause length and the presence or absence of fillers (Brennan & Williams, 1995).

Another potential cue of respondent uncertainty might be *hedges*: approximators like *about*, as in "We have about four bedrooms," and shields like *I think*, as in "I think we have four bedrooms" (see Prince, Frader, & Bosk, 1982., for a taxonomy of hedges). Speakers have been shown to hedge more when they have more potential ways to describe objects (Brennan & Clark, 1996); certainly the common interpretation of hedges is that they denote speaker uncertainty.

A third set of potential cues of respondent uncertainty are *discourse markers* like *well* (as in "Well, yes I did buy some furniture") and *oh* (as in "Oh, I did buy some furniture"). Some discourse markers, it has been argued, alert listeners that what comes next is unexpected, or that the speaker isn't entirely sure (Schiffrin, 1994).

The final potential cue of respondent uncertainty we examine in this study is what Drew (1984) called *reporting*, in which one answers a question in a way that leaves the responsibility of answering to the person who posed the question. For example, if person A asks person B "Do you like punk rock?" and B responds "I like The Clash," B has left it up to A to decide whether The Clash's music counts as punk rock. A survey analog of this would be answering the household furniture question with "I bought a floor lamp."

METHOD

Our strategy was to examine whether these various linguistic cues occur reliably more often in situations where respondents have been shown to be likely to misunderstand survey concepts than those in which they have been shown to understand well. We thus contrast the use of these potential cues in (1) straightforward situations where (for example) a respondent who purchased an end table or a sofa (objects that clearly count as furniture) is asked about household furniture purchases with (2) situations like our floor-lamp-purchasing respondent's, in which her circumstances map onto the question concepts in a more complicated way.

We coded a corpus of 42 transcripts of telephone "survey interviews" from the Schober and Conrad (1997) laboratory study. The survey included 12 questions from three different ongoing government surveys. Four questions, adapted from the Consumer Price Index Housing survey, were about housing (e.g., "How many people live in this house?"). Four

questions, adapted from the Current Point of Purchase Survey, were about purchases (e.g., “Last year, did Carla purchase or have expenses for car tires?”). Four questions, adapted from the Current Population Survey, were about employment (e.g., “Last week did Pat have more than one job, including part-time, evening or weekend work?”).

In the interviews, respondents answered these questions based upon fictional scenarios. For each question there were two alternate scenarios, one of which mapped onto the question in a straightforward way (e.g., a purchase receipt for an end table for the household furniture question) and the other of which mapped onto the question in a complicated way (e.g., a purchase receipt for a floor lamp). For each respondent half of the scenarios had straightforward mappings and the other half had complicated mappings. Respondents nearly always gave correct answers for straightforward mappings, but they were far more likely to give inaccurate answers for the complicated mappings.

Half of the respondents in the transcribed interviews participated in strictly standardized interviews, in which interviewers left the interpretation of questions entirely up to the respondents. The other half participated in “conversational” interviews in which interviewers tried to make sure that respondents had interpreted the questions according to the official definitions, saying whatever it took to make this happen. In these interviews respondents were instructed to ask for clarification whenever they had the slightest doubt about how to interpret survey questions; but conversational interviewers not only provided clarification when respondents requested it, but also whenever they believed that respondents might need it in order to interpret the questions correctly.

We coded for each potential uncertainty cue—pauses, fillers, repairs, hedges, discourse markers, and reporting—in the respondent utterance that immediately followed the interviewer's question. To verify whether the coding was reliable, two judges coded 4 of the 42 transcripts (2 standardized and 2 conversational interviews). Reliability of the coding was satisfactory, averaging .75 (Cohen's kappa), ranging from .68 for repairs to .92 for pauses. Because so few discourse markers appeared in the corpus, we dropped them from further consideration.

We then asked two main questions. First, do respondents use potential cues more when mappings are complicated? If so, this would suggest that these linguistic cues provide reasonable evidence for interviewers (or even computerized interviewing systems) that respondents may need clarification.

Second, does the use of these cues predict whether respondents are likely to provide answers that match the official definitions? Even if these cues reliably predict complicated mappings, they might not predict ultimate inaccuracy, as respondents may be confident about incorrect answers or uncertain about correct answers.

RESULTS

Explicit requests for clarification. Of course, one should expect that the very best predictors that respondents may misunderstand survey questions are respondents' explicit requests for clarification in the first turn immediately following the interviewer's question. As Table 1 shows, respondents did this rarely in standardized interviews (no doubt because such requests would not be honored), and reliably more often in conversational interviews, $F(1,40) = 16.00, p < .001$, as they had been instructed (and where such requests were more likely to be effective). As the table also shows, respondents explicitly requested clarification much more often for complicated cases than for straightforward cases, $F(1,40) = 21.78, p < .001$. This demonstrates that explicit requests for clarification are indeed excellent markers of respondents' likelihood of misunderstanding survey questions.

	Straightforward situations	Complicated situations
Standardized interviews	0%	2%
Conversational interviews	16%	34%

Table 1: Percentage of questions for which respondents explicitly requested clarification in first turn after question

Pauses. The mere presence of a pause longer than one second in the first turn following the interviewer's question was not a good predictor that respondents were likely to misunderstand the question. As Table 2 shows, although respondents paused reliably more often for complicated mappings than for straightforward mappings in both types of interviews, $F(1,40) = 21.36, p < .001$, respondents paused so often for straightforward cases that this cue isn't very informative.

In the standardized interviews, the length of respondents' pauses was a more reliable marker of respondent uncertainty; respondents paused longer in their first utterance after hearing the question for complicated mappings (2.2 seconds) than for

	Straightforward situations	Complicated situations
Standardized interviews	51%	67%
Conversational interviews	58%	74%

Table 2: Percentage of questions for which respondents paused longer than one second in first turn after question

straightforward mappings (1.4 seconds) in the standardized interviews, $F(1,20) = 15.68, p < .001$. But the opposite was true in the conversational interviews: respondents paused longer for straightforward mappings (2.4 seconds) than for complicated mappings (2.0 seconds), $F(1,40) = 6.05, p < .02$.

Fillers. Overall, fillers were good predictors of respondent uncertainty. Respondents were more likely to produce fillers for complicated mappings than for straightforward mappings, $F(1,40) = 7.52, p < .01$. But as Table 3 shows, this was largely because of the predictive value of fillers in standardized interviews, where the difference in rates of fillers was much greater than in conversational interviews, interaction of mappings x interview type, $F(1,40) = 3.94, p = .054$. In the standardized interviews, respondents were almost three times more likely to produce a filler in their first turn following the question for complicated mappings (33% of cases) than for straightforward mappings (13%).

	Straightforward situations	Complicated situations
Standardized interviews	13%	33%
Conversational interviews	48%	51%

Table 3: Percentage of questions for which respondents produced at least one filler in first turn after question

Repairs. As Table 4 shows, in both types of interviews respondents were three times more likely to repair their first utterance following the question for complicated mappings than for straightforward mappings, $F(1,40) = 20.21, p < .001$. Even though repairs didn't happen so often in the standardized interviews, they were still very good predictors of when respondents might be likely to misinterpret questions.

	Straightforward situations	Complicated situations
Standardized interviews	2%	8%
Conversational interviews	13%	31%

Table 4: Percentage of questions for which respondents produced at least one repair in first turn after question

Hedges. Hedging in the first turn immediately after the question was rare in this corpus (23 cases in all), and the pattern of hedging was inconclusive. Respondents in the standardized interviews hedged slightly more often for complicated than for straightforward mappings, $F(1,20) = 4.14, p < .06$. But respondents in the conversational interviews hedged more often for straightforward mappings than for complicated mappings, $F(1,40) = 10.76, p < .01$. This is hard to explain; perhaps respondents in conversational interviews who weren't sure about answers that should have been obvious were more likely to mark those answers with hedges, as a way of reducing embarrassment.

Reporting. Reporting was rare in the first turn after a question was asked, even though reporting at some point was fairly common (see Schober and Conrad, 1997). As Table 5 shows, respondents reported reliably more often in conversational interviews than in standardized interviews, $F(1,40) = 4.51, p < .05$. Although there was no overall effect of mapping on how often respondents reported, in standardized interviews reporting appeared to be a perfect predictor of complicated mappings; respondents only reported when faced with a complicated mapping, $F(1,20) = 8.00, p < .01$.

	Straightforward situations	Complicated situations
Standardized interviews	0%	5%
Conversational interviews	5%	9%

Table 5: Percentage of questions for which respondents reported their circumstances, rather than answering, in first turn after question

Multiple cues. We should note that sometimes several cues co-occurred in the same utterances, as in this example from the corpus (cues in italics):

Interviewer: Has Kelly purchased or had expenses for household furniture.

Respondent: (3.5 sec) *Um h-* household furniture or are we (1.7 sec) *uh* referrin' to *uh* fix-fixtures and accessories?

Further analyses will allow us to determine whether respondent utterances with multiple cues were better predictors of complicated mappings than utterances with single cues.

Cues and response accuracy. Several of the potential cues we examined did indeed prove to be reliable cues of complicated mappings—of cases where respondents might be liable to misinterpret questions. To what extent did these cues predict the accuracy of respondents' ultimate answers? Note that this question is a different one for standardized interviews, where the interpretation of questions is left up to respondents, than for conversational interviews, where respondent cues of uncertainty might lead sensitive interviewers to provide clarification.

Because respondents were almost perfectly accurate for straightforward mappings, we focus on complicated mappings only; also, we focus on pauses and fillers, because they happened sufficiently often that we could examine response accuracy. As Table 6 shows, in both standardized and conversational interviews respondents were more likely to be accurate when they paused than when they did not, $F(1,28) = 9.66, p < .01$. The findings are virtually identical for fillers, $F(1,24) = 9.48, p < .01$. Of course, in standardized interviews response accuracy was generally poor; the accuracy in cases where respondents had paused was still under 40%. In any case, pauses and fillers certainly do not seem to be direct indicators of respondent misinterpretation.

	Respondent paused	Respondent didn't pause
Standardized interviews	39%	15%
Conversational interviews	88%	75%

Table 6: Ultimate response accuracy for complicated situations when respondents paused in first turn after question vs. when they didn't

We interpret this to mean that even though pauses and fillers may predict complicated mappings, they may not directly predict inaccurate responding. In indicating their uncertainty by pausing or *umming*, respondents may actually be demonstrating that they are *more* likely to be recognizing the potential

problem in mapping their circumstances onto the question. Respondents who do not provide cues of uncertainty may be confident about their inaccurate responses.

SUMMARY AND CONCLUSIONS

Our findings show that some ordinary linguistic and paralinguistic behaviors can reliably indicate survey respondents' likelihood of having circumstances that do not map onto questions in a straightforward way. In particular, respondent repairs and reporting are good indicators, and for standardized interviews, so are pause length and *ums* and *uhs*. Other potential indicators—hedges and discourse markers—were not particularly good indicators, at least in this corpus.

Our findings also show that interviewing style affects how often respondents produce these cues of uncertainty: respondents *ummed*, repaired and reported more often in conversational interviews. Perhaps respondents recognized that uncertainty displays were more likely to elicit help in conversational interviews. Or perhaps respondents worked harder in conversational interviews, because they would be held accountable for inaccurate responses, and so they displayed their greater cognitive effort more with these cues.

Although many of these cues predicted complicated mappings (and thus, we argue, respondent uncertainty), they did not predict inaccurate answers. In fact, for complicated mappings, fluent responses were more likely to be wrong. Perhaps respondents who produce fillers, repairs, etc. at some level recognize that their circumstances are complicated; respondents who don't produce them may either be more confident in their answers, or fail to recognize that their conceptions might differ from the survey designers'.

So are uncertainty cues useful for survey researchers to pay attention to? We argue that they are, for several reasons. As Mathiowetz (1998) has proposed, uncertainty expressions may be a good data source for imputation models. In our study, at least some discourse cues of uncertainty predict complicated mappings, which in general are more likely to be inaccurate³. Given how rarely survey respondents ask for clarification when they really need it, these sorts of cues might allow interviewers to provide clarification more appropriately—that is,

³ Note that the slight variation in accuracy for complicated mappings when cues were present or absent is negligible compared to the overall benefit for providing clarification for complicated mappings.

when respondents need it. (We suspect that some interviewers already rely on just these sorts of cues when they clarify questions without being explicitly asked by respondents.) We propose that computerized interviewing systems of the future (e.g., Bloom, 1999; Schober, Conrad, & Bloom, 1999) might be able to use the information that these cues convey to choose when to provide users with unsolicited clarification.

But we should note that the absence of uncertainty cues does *not* mean that respondents understand questions as intended—only that respondents *believe* that they do. Respondents may not take the time or have the metacognitive resources to notice that their own conceptualizations for ordinary words like “furniture” or “work” could differ from survey designers’ conceptions. It isn’t clear how survey researchers can improve response accuracy when respondents are confident of their incorrect interpretations of survey questions.

REFERENCES

- Bloom, J.E. (1999). *Linguistic markers of respondent uncertainty during computer-administered survey interviews*. Ph.D. Dissertation, New School for Social Research.
- Bortfeld, H., Leon, S.D., Bloom, J.E., Schober, M.F., & Brennan, S.E. (2000). Which speakers are most disfluent, and when? Manuscript under review.
- Brennan, S. E., & Clark, H.H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22, 482-1493.
- Brennan, S.E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, 34, 383-398.
- Conrad, F.G. (1999). Customizing survey procedures to reduce measurement error. In M.G. Sirken, D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur, & R. Tourangeau (eds.), *Cognition and survey research* (pp. 77-93). New York: Wiley.
- Drew, P. (1984). Speakers' reportings in invitation sequences. In J.M. Atkinson & J. Heritage (eds.), *Structures of social action: Studies in conversation analysis* (pp. 129-151). New York: Cambridge University Press.
- Fowler, F.J., & Mangione, T.W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error*. Newbury Park, CA: SAGE Publications, Inc.
- Graesser, A.C., & McMahan, C.L. (1993). Anomalous information triggers questions when adults solve quantitative problems and comprehend stories. *Journal of Educational Psychology*, 85(1), 136-151.
- Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a standard maximum silence of approximately one second in conversation. In D. Roger & P. Bull (eds.), *Conversation* (pp. 166-196).. Clevedon: Multilingual Matters,
- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Mathiowetz, N.A. (1998). Respondent expressions of uncertainty: Data source for imputation. *Public Opinion Quarterly*, 62, 47-56.
- Prince, E.F., Frader, J., & Bosk, C. (1982). On hedging in physician-physician discourse. In R.J. diPietro (ed.), *Linguistics and the professions* (pp. 83-97). Hillsdale, NJ: Ablex.
- Schaeffer, N.C., Maynard, D.W., & Craddock, R. (1993). *Negotiating certainty: Uncertainty proposals and their disposal in standardized survey interviews*. University of Wisconsin, Madison, Center for Demography and Ecology Working Paper 93-25.
- Schiffrin, D. (1994). *Discourse markers*. Cambridge, UK: Cambridge University Press.
- Schober, Michael F. (1999). Making sense of questions: An interactional approach. In M.G. Sirken, D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur, & R. Tourangeau (eds.), *Cognition and survey research* (pp. 77-93). New York: Wiley.
- Schober, M.F., & Conrad, F.G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, 60, 576-602.
- Schober, M.F., & Conrad, F.G. (2000). A collaborative view of standardized survey interviews. In D. Maynard, H. Houtkoop, N.C. Schaeffer, & J. van der Zouwen (eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview*. New York: Wiley, in press.
- Schober, M.F., Conrad, F.G., & Bloom, J.E. (1999). Enhancing collaboration in computer-administered survey interviews. In *Proceedings of the AAAI Fall Symposium "Psychological Models of Communication in Collaborative Systems"* (pp. 108-115). Menlo Park, CA: AAAI Press.
- Smith, V.L., & Clark, H.H. (1993). On the course of answering questions. *Journal of Memory and Language*, 32, 25-38.